多粒度时序特征在离网预测中的应用 *

石鸿斌，严建峰，白瑞瑞，徐彩旭，徐广根

(苏州大学 计算机科学与技术学院，江苏 苏州 215006)

摘　要：电信运营商为了发现可能离网的客户，针对不同的场景研究开发了多种离网预测模型。目前的离网预测模型首先选择一种时间粒度抽取特征，之后使用机器学习算法对抽取的数据建模。这类方法只考虑了模型对分类性能的影响，没有充分考虑数据的作用。针对上述问题，提出一种使用多种时间粒度抽取特征的方法，并尝试在模型训练的不同阶段对不同粒度的特征进行融合。实验结果表明，使用多种粒度抽取特征训练出来的模型性能会明显优于使用单一粒度抽取特征的模型。

关键词：离网预测；时序数据；多粒度

中图分类号：TP391

Application of multi-grain temporal features in churn prediction

Shi Hongbin, Yan Jianfeng, Bai Ruirui, Xu Caixu, Xu Guanggen

(*School of Computer &Technology*, *Soochow University*, *Suzhou Jiangsu* 215006, *China*)

Abstract: Telecom operators have developed multiple churn prediction models to find potential users for different scenes. The present churn prediction models firstly select a kind of time granularity to extract features, then model the extracted data using machine learning algorithm. Such approaches only consider the influence of the model on classification performance, but the role of data is not fully considered. To solve this problem, this paper proposed a method which extracts multi-grain temporal features, and try to integrate different granularity features at different training phases. Experimental results show that the performance of the model trained with multi-grain features is obviously superior than that trained with single granularity features.

Key Words: churn prediction; time series data; multi-granularity

0　引言

近年来，用户流失问题成为电信运营商亟需解决的问题，同时也引起了业界的广泛关注和研究。图 1 统计了上海市某运营商 2G/3G 预付费用户在 2015 年 9 月到 2016 年 8 月期间的离网率。其中活跃用户指的是每个月的通话总时长、上网总流量或者发送短信总次数超过阈值的用户，可以给运营商带来更大的商业利润，非活跃用户指的是通话总时长等三项指标没有达到阈值的用户，给运营商带来的利润较少。

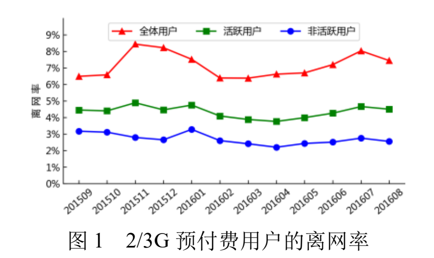


图 1　2/3G 预付费用户的离网率

从图 2 可以看出，活跃用户占比较高，约为 63%。由于非活跃用户的离网可以通过阈值较精确的预测出，并且对运营商的利润影响不大，因此聚焦活跃用户的离网意义重大。研究表明，成功发展一个新客户的成本是挽留现有客户的 3 倍以上，而采用合理的维系挽留策略挽留客户的成功率高于发展新用户的成功率[1-3]。本文针对活跃用户群体，研究和实现了基于多粒度时序特征的离网预测模型，为运营商采取有针对性的维系挽留策略提供了科学依据，最大程度挽留用户并增加利润。

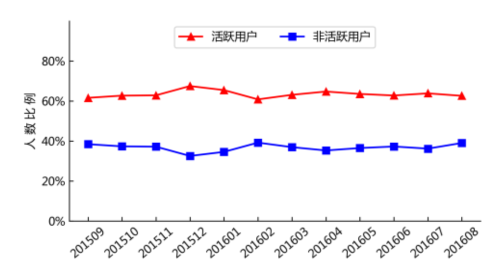


图 2　2/3G 活跃用户和非活跃用户占总人数比例

基金项目：国家自然科学基金资助项目（61572339）；江苏省科技支撑计划重点项目（BE2014005）

作者简介：石鸿斌（1992-），男，江苏昆山人，硕士研究生，主要研究方向为机器学习与数据挖掘（shb_6600@qq.com）；严建峰（1978-），男，副教授，博士研究生，主要研究方向为机器学习；白瑞瑞（1989-），女，博士研究生，主要研究方向为机器学习与数据挖掘；徐彩旭（1993-），男，硕士研究生，主要研究方向为机器学习与数据挖掘；徐广根（1992-），男，硕士研究生，主要研究方向为机器学习与数据挖掘．

从运营商的角度，用户有两种状态，在网和离网。用户能给运营商带来收入，这个用户就被认为是在网。当用户不能给运营商带来任何收入时，这个用户就被认为是离网。离网预测是通过分析用户的历史行为数据，从中找到一些行为规律，从而对用户未来的行为进行预测。目前国内外的离网预测研究主要从三个方面提高性能：构建有用的特征、构造好的分类器和分类器集成[4,5]。构建特征阶段主要使用某种时间粒度抽取用户的个人信息，行为数据（包括通话，短信、上网等）特征，例如按月[6,7]，按日[8]或按 2 小时[9]。月粒度指的是将时间按月划分，统计用户在每个月的数据，如话费、通话量、短信量等。日粒度和小时粒度与月粒度类似。构建分类器阶段常见的算法包括神经网络(neural network)[6-9]，逻辑回归(logistic regression)[10~12]，支持向量机(support vector machine)[13 -15]，决策树(decision tree)[14~17]，K 近邻(K-nearest neighbors)[18]等。其中文献[6]通过神经网络算法构造离网预测模型，文中构造了三种网络结构，包括:具有两层全连接层的结构；具有三层全连接层的结构；拥有一个卷积-池化层和两层全连接层的结构。实验发现，集成学习被认为是提高模型性能的一种有效的方法，在离网预测问题中有支持向量机集成[19-21]，决策树集成[22]等。文献[19]训练多个不同的支持向量机并融合，文献[22]训练多个不同的决策树并融合。综上所述，大多数工作集中在设计分类器和模型融合方面，而在特征抽取阶段，通常只选取单一的时间粒度特征，很少考虑使用多个时间粒度同时抽取特征[23]。

针对上述问题，本文提出了一种基于多种时间粒度特征融合的离网预测方法。在模型训练阶段，按照特征值是否随着时间变化，分为不变特征和变化特征，又对变化的特征进行月粒度和日粒度区分，通过不同组合方式形成不同的数据集，分别采用随机森林和 GBDT 训练得到六个模型。在模型融合阶段，采用了平均法和 Stacking 方法进行模型融合，通过融合不同的模型，最后将 Top 25000 的查准率提高到 0.5696。

1　相关理论基础

1.1　算法

1.1.1 随机森林

随机森林是由 Breiman 在 2001 年[24]提出的一种算法，是 Bagging 的一个变体。Bagging 在构造样本的时候，从数据集 D 中有放回的抽取一个样本，抽取 m 次，得到一个包含 m 个样本的数据集 D'，重复上述操作得到 T 个数据集，在每个数据集上训练一个基分类器，最后使用投票的方式融合多个基学习器。随机森林在 Bagging 的基础上加入了随机属性划分。假设样本中有 d 个属性，对于随机森林中的基决策树，选取样本中的 k 个属性用来划分叶子节点。通常选取 $k=\log_2 d$ 。随机森林通过同时加入样本扰动和属性扰动，显著提高了基学习器的泛化性能。

1.1.2 GBDT

GBDT 与随机森林一样训练多棵决策树，在多棵决策树的基础上得到结果。随机森林在训练的过程中，各棵树之间可以同时训练，互不依赖。与随机森林不同的是，GBDT 在训练第 k 棵树时，依赖于前 $k-1$ 棵树的输出。假设 f_k 表示 GBDT 中第 i 个样本的预测值，则 $\hat{y}_i$ 可以表示为

$$\hat{y}_i=\sum_{k=1}^{K} f_k(x_i) \tag{1}$$

其中：K 是树的棵树，f_k 表示第 k 棵树的输出。对于 GBDT，本文有以下目标函数：

$$\begin{aligned} obj^{(t)} &= \sum_{i=1}^{n} l(y_i,\hat{y}_i^{(t)})+\sum_{i=1}^{t}\Omega(f_i) \\ &= \sum_{i=1}^{n} l(y_i,\hat{y}_i^{(t-1)}+f_t(x_i))+\Omega(f_t)+constant \end{aligned} \tag{2}$$

其中：l 是损失函数，Ω 是正则项。将损失函数在 $\hat{y}$ 处用泰勒公式展开得

$$\begin{aligned} obj^{(t)} = &\sum_{i=1}^{n}\left[l(y_i,\hat{y}_i^{(t-1)})+g_i f_t(x_i)+\frac{1}{2}h_i f_t^2(x_i)\right] \\ &+\Omega(f_t)+constant \end{aligned} \tag{3}$$

其中 g_i 和 h_i 被定义为

$$\begin{aligned} g_i &= \partial_{\hat{y}_i^{(t-1)}} l(y_i,\hat{y}_i^{(t-1)}) \\ h_i &= \partial^2_{\hat{y}_i^{(t-1)}} l(y_i,\hat{y}_i^{(t-1)}) \end{aligned} \tag{4}$$

由泰勒公式展开可知，训练第 t 棵树时依赖前 $t-1$ 棵树的输出。

1.2　模型融合方法

模型融合通过组合多个单模型，进一步提升模型的性能。假设 $h_i(x)$ 表示第 i 个模型的输出，T 表示模型的数量，本文使用两种模型融合方法：

1.2.1 Averaging

模型的最终输出为

$$H(x)=\frac{1}{T}\sum_{i=1}^{T} h_i(x) \tag{5}$$

即得到每个模型的预测概率后，将这些概率取平均值作为最终输出。

1.2.2 Stacking

Stacking 的步骤如下所示：

算法：Stacking

输入：训练集 D

基学习算法 $f_1, f_2,\ldots, f_T$，

次级学习算法 f

输出：$H(x)=h(h_1(x),h_2(x),\ldots,h_T(x))$

将训练集 D 分为 D_1 和 D_2

使用 D_1 和 $f_1, f_2,\ldots, f_T$ 训练模型 $h_1,h_2,\ldots,h_T$

将预测值 $h_1(D_2),h_2(D_2),\ldots,h_T(D_2)$ 组合成 D'

使用 D' 和 f 训练模型 h

本文使用逻辑回归作为次级学习算法。在预测阶段使用基学习器对数据集进行预测得到预测结果，将预测结果作为次级学习器的输入，次级学习器的预测结果作为最终预测结果。

2　基于多粒度时序特征的离网预测模型

本文所提的离网预测模型框架由四个部分组成，分别是数据准备、特征抽取、模型训练和模型融合。模型框架如图 3 所示。为了验证提出的框架的通用性，在模型训练和模型融合阶段分别尝试不同的方法。

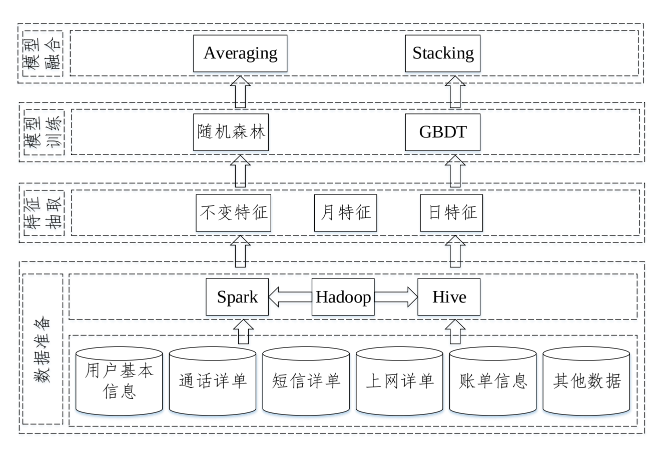


图 3　离网预测模型框架

2.1　数据准备

实验所需的数据来自于上海某运营商的大数据平台，大数据平台每天会产生约 2.3 TB 的数据，这些数据包括 BSS (business support system)数据以及 OSS (operation support system)数据。BSS 又称为业务支持系统，数据包括用户基本信息、用户行为、账单信息、语音数据、短信数据及通话详单等，每天产生大约 24GB 的数据。目前，BSS 支持的用户包括 2G/3G/4G 预付费用户和后付费用户。本文以 2G/3G 预付费活跃用户为研究对象，从三十多张表中选出 10 张表作为数据源，包括用户基本信息日表、用户基本信息月表、用户行为日表、套餐表、用户余额月表、用户语音通话详单表、用户短信收发详单表、账单月表、终端表和充值表。实验中使用的训练集共有 1 415 429 条记录，测试集共有 1 380 154 条记录。

2.2　特征抽取

本文将特征分为三种，不变特征、月特征和日特征。不变特征指的是特征的值不随时间的变化而变化，或者随时间线性变化。本文从用户基本信息中抽取部分不变特征，比如性别、年龄等。月特征指以月为单位抽取出来的数据，如每月的话费、通话时长、上网流量等。日特征指以日为单位抽取出来的数据，如每天的话费、通话时长、上网流量等。在数据库中数据有月和日两种时间粒度，在抽取月特征时，对于粒度为月的数据，直接取出对应的值，对于粒度为日的数据，求该特征的在某月的平均值作为月特征。本文共抽取不变特征 3 个，月特征每月 52 个，日特征每日 34 个。表 2 为抽取出来的部分特征样例。

表 2　特征样例

	特征名称	特征含义	特征类型
不变特征	is_shanghai	是否是上海人	category
	sex	性别	category
	age	年龄	int
	terminal	终端	category
月特征	ld_call_dur	长途主叫时长	float
	fest_call_dur	节假日通话时长	float
	sms_bill_cnt	短信计费量	float
	…	…	…
	gift_flex_value	套餐内流量	float
日特征	caller_cnt	主叫次数	float
	feat_call_dur	节假日通话时长	float
	sms_bill_cnt	短信计费量	float
	balance	余额	float
	…	…	…
	serve_sms_count	服务类短信条数	float

数据集使用滑动窗口的方式组合，各个特征的时间窗口如表 3 所示。训练集使用的是 2015 年 1 月到 2015 年 12 月共 12 个月的月粒度特征以及 2015 年 10 月 1 日到 2015 年 12 月 31 日共 92 天的日粒度特征组成，标签使用 2016 年 1 月的用户离网和在网标签。测试集使用的是 2015 年 2 月到 2016 年 1 月共 12 个月的月粒度特征以及 2015 年 11 月 1 日到 2016 年 1 月 31 日共 92 天的日粒度特征组成，标签使用 2016 年 2 月的用户离网和在网标签。日粒度特征粒度较细，可以较精确的描述用户近期的行为，但天数取太多容易导致特征过多难以训练。月粒度特征粒度较粗，特征较少，适合用来记录用户长期的行为趋势。使用 12 个月的月粒度特征和 92 天的日粒度特征可以同时发挥两种特征的优势且不加入太多特征。

假设 F_1、 F_2 和 F_3 分别表示不变特征，月特征和日特征，如图 4 所示，通过拼接不变特征和月特征得到数据集 D_1，拼接不变特征和日特征得到数据集 D_2，拼接不变特征、月特征和日特征得到数据集 D_3。

表 3　数据集时间窗口

训练集	不变特征	
	月特征	201501~201512
	日特征	20151001~20151231
	标签	201601
测试集	不变特征	
	月特征	201502~201601
	日特征	20151101~20160131
	标签	201602

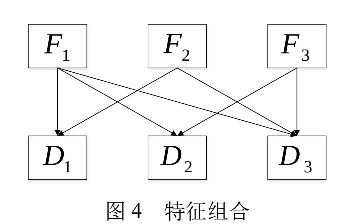


图 4　特征组合

2.3　模型训练

树型模型是一类在工业界中非常常用的模型，文献[14~17]在离网预测任务中使用树型模型，均取得了比较好的效果。同时树型模型可以在一定程度上做特征选择，降低不同时间粒度的特征融合时出现冗余的风险。在树型模型中随机森林和GBDT 是两种效果非常好的模型，因此本文在模型训练阶段使用这两种模型进行训练。

在上述三个数据集上分别使用随机森林和GBDT 训练得到6 个模型 RF_1、RF_2、RF_3、$GBDT_1$、$GBDT_2$、$GBDT_3$。为了保证公平，分别对三个随机森林和三个 GBDT 使用相同的参数。对于随机森林，本文设置树棵树为 200 棵，对于 GBDT，设置学习率为 0.1，树最大深度为 8，L2 正则为 50。其他参数使用工具包默认值。

2.4　模型融合

在模型融合方法中 Stacking 是一种通用且有效的方法，文献[25]在离网预测任务中使用了 Stacking。同时 Averaging 是另一种比较简单有效的融合方法，在本文中同时使用 Averaging 和 Stacking 做对比实验。

假设 A 表示使用 Averaging 融合模型，S 表示使用 Stacking 融合模型，对于模型训练阶段训练得到的 6 个模型，实验从 1)月粒度特征和日粒度特征融合；2)分类器融合两个角度尝试了14 种融合方式，其中 R_3 和 G_3 尝试了两种粒度的特征直接拼接后训练单个分类器，分类器融合又分为同种分类器的融合和不同分类器的融合。当模型融合方式为 Stacking 时，本文使用训练集中 90%的数据训练基学习器，10%的数据训练次级学习器。具体融合方式如下：

1)月粒度特征和日粒度特征融合

a)将 RF_3 的预测结果作为最终输出(R_3)；

b)将 $GBDT_3$ 的预测结果作为最终输出(G_3)。

2)平均法融合同种分类器

a)将 RF_1 和 RF_2 的预测值用 Averaging 融合（$R_{12}+A$)；

b) 将 RF_1、RF_2 和 RF_3 的预测值用 Averaging 融合($R_{123}+A$)；

c) 将 $GBDT_1$ 和 $GBDT_2$ 的预测值用 Averaging 融合($G_{12}+A$)；

d)将 $GBDT_1$、$GBDT_2$ 和 $GBDT_3$ 的预测值用 Averaging 融合($G_{123}+A$)。

3）Stacking 融合同种分类器

a)将 RF_1 和 RF_2 的预测值用 Stacking 融合($R_{12}+S$)；

b)将 RF_1、RF_2 和 RF_3 的预测值用Stacking融合($R_{123}+S$)；

c)将 $GBDT_1$ 和 $GBDT_2$ 的预测值用 Stacking 融合($G_{12}+S$)；

d)将 $GBDT_1$、$GBDT_2$ 和 $GBDT_3$ 的预测值用 Stacking 融合($G_{123}+S$)。

4）平均法融合不同分类器

a)将 RF_1、RF_2、$GBDT_1$ 和 $GBDT_2$ 的预测值用 Averaging 融合($R_{12}G_{12}+A$)；

b) 将所有 6 个模型的预测值用 Averaging 融合($R_{123}G_{123}+A$)。

5）Stacking 融合不同分类器

a)将 RF_1、RF_2、$GBDT_1$ 和 $GBDT_2$ 的预测值用 Stacking 融合($R_{12}G_{12}+S$)；

b)将所有 6 个模型的预测值用 Stacking 融合($R_{123}G_{123}+S$)。

3　实验结果和分析

本文实验基于运营商的大数据平台，平台共有 15 个节点，每个节点有 24 个 CPU 核心和 188 GB 内存。在数据准备和特征抽取阶段使用 Spark 在所有节点上计算，模型训练和模型融合阶段使用 Scikit-learn 和 XGBoost 在单节点计算。

3.1　评价标准

实验中使用查准率(precision)、查全率(recall)和 AUC 作为评价指标。由于在离网预测模型中预测正确比预测全更重要，因此主要考察查准率，查全率和 AUC 作为参考。分类结果的混淆矩阵如表 4 所示。

表 4　混淆矩阵

真实情况	预测结果	
	正例	反例
正例	TP(真正例)	FN(假反例)
反例	FP(假正例)	TN(真反例)

查准率 P 和查全率 R 的定义分别为

$$P=\frac{TP}{TP+FP}\qquad R=\frac{TP}{TP+FN}\tag{6}$$

实验中取 Top25000 的查准率和查全率。首先将预测得到的概率值降序排序，将前 25000 个预测样本标为正例，剩下的标为负例，统计样本的查准率和查全率。

AUC 是 ROC 曲线下的面积。ROC 曲线的纵轴是真正例率(TPR)，横轴是假正例率(FPR)，定义分别为

$$TPR=\frac{TP}{TP+FN}\qquad FPR=\frac{FP}{TN+FP}\tag{7}$$

将预测得到的概率值降序排序，全部标为反例，在坐标(0, 0)标记点。依次将概率值最大的样本标记为正例，重新计算

TPR 和 *FPR* 并在坐标中标记。连接这些标记点得到 ROC 曲线。

3.2 实验结果

使用随机森林和 GBDT 分别在 D_1 和 D_2 数据集上训练模型并预测，得到 R_1 、R_2 、G_1 、G_2 四个模型，然后将同一种时间粒度训练的模型用平均法或 Stacking 融合作为 Baseline，与 3.4 中提到的 14 种融合方法作对比。表 5~10 是实验结果。

表 5 Baseline 性能

模型	P	R	AUC
R_1	0.5147	0.2554	0.8815
R_2	0.5072	0.2517	0.8583
G_1	0.4628	0.2297	0.8879
G_2	0.4801	0.2382	0.8899
R_1G_1+A	0.5286	0.2623	0.8957
R_1G_1+S	0.5342	0.2651	0.8952
R_2G_2+A	0.5142	0.2552	0.8834
R_2G_2+S	0.5068	0.2515	0.8859

从表 5 中可以看到，Top25000 时，随机森林的查准率和查全率比 GBDT 好，而 GBDT 的 AUC 优于随机森林。此外，随机森林在月粒度特征上性能较好，而 GBDT 在日粒度特征上性能更好。在对使用单个时间粒度训练的模型融合时候，月粒度特征比日粒度特征有更好的表现。

表 6 月粒度特征和日粒度特征融合

模型	P	R	AUC
R_3	0.5454	0.2707	0.8787
G_3	0.5159	0.2560	0.9028

从表 6 中可以看到，将月粒度特征和日粒度特征直接拼接训练模型后，R_3 和 G_3 的查准率和查全率都有很明显的提升，G_3 的 AUC 也有小幅度提升。R_3 在没有进行多模型融合的情况下查准率超过了 Baseline 中的 4 种多模型融合方法，G_3 也超过了 2 种融合方法，说明直接拼接月和日粒度特征对模型效果有很明显的提升。

表 7 平均法融合同种分类器

模型	P	R	AUC
$R_{12}+A$	0.5507	0.2733	0.8871
$R_{123}+A$	0.5545	0.2752	0.8876
$G_{12}+A$	0.5222	0.2591	0.9014
$G_{123}+A$	0.5449	0.2704	0.9059

在表 7 中比较了使用平均法融合不同模型的结果，可以看到对模型 R_1 和 R_2 取平均值后的性能要高于单个模型的性能，查准率从 R_1 的 0.5174 和 R_2 的 0.5072 提高到 0.5507，入 R_3 后可以进一步提高性能，达到 0.5545，和 Baseline 中的 2 个平均法融合模型的方法比也同样有提高，从 0.5286 和 0.5142 提高到 0.5507 并且在加入 R_3 后，模型的查准率从 0.5507 提高到了 0.5545。查全率和 AUC 总体持平。对于 GBDT，和 Baseline 的 2 个平均法融合模型的方法比总体持平，提升没有随机森林明显。这说明将具有差异性的基分类器通过平均法融合后，可以减少泛化误差，提高模型性能。

表 8 Stacking 融合同种分类器

模型	P	R	AUC
$R_{12}+S$	0.5487	0.2723	0.8883
$R_{123}+S$	0.5585	0.2772	0.8895
$G_{12}+S$	0.5014	0.2488	0.8995
$G_{123}+S$	0.5246	0.2604	0.9045

表 8 与表 7 类似，表 8 将表 7 中的平均法替换为 Stacking。可以看到，使用 Stacking 融合随机森林时同样可以提高模型性能，融合 GBDT 时效果稍差。对于相同的基分类器，采用集成的思想，包括平均法和 Stacking，都会使融合后模型的泛化性能高于单个分类器的泛化性能。

表 9 平均法融合不同分类器

模型	P	R	AUC
$R_{12}G_{12}+A$	0.5622	0.2790	0.9027
$R_{123}G_{123}+A$	0.5696	0.2872	0.9039

表 10 Stacking 融合不同分类器

模型	P	R	AUC
$R_{12}G_{12}+S$	0.5564	0.2761	0.9048
$R_{123}G_{123}+S$	0.5624	0.2791	0.9068

表 9 和 10 展示了对不同类型的分类器进行融合的结果。可以看出将 R_{123} 和 G_{123} 通过平均法融合后，查准率从 $R_{123}+A$ 的 0.5545 和 $G_{123}+A$ 的 0.5449 提高到 0.5696，通过 Stacking 融合后查准率从 $R_{123}+S$ 的 0.5585 和 $G_{123}+S$ 的 0.5246 提高到 0.5624。在查全率和 AUC 上同样有提升。可以说明，对有差异的不同类型的分类器融合之后可以得到更好的性能。并且平均法融合比 Stacking 有更好的表现。

3.2.1 分类器数量对性能的影响

从表 5~10 中可以看出，模型最终性能随着分类器数量的增加提高。以 GBDT+简单平均法为例，如图 5 所示，$G_{12}+A$ 的 Top25000 的查准率、查全率和 AUC 分别比只使用 G_1 提高了 0.0594、0.0294 和 0.0135，加入 G_3 后的 $G_{123}+A$ 比只使用 G_1 提高了 0.0821、0.0407 和 0.018，$R_{12}G_{12}+A$ 比只使用 G_1 提高了 0.0994、0.0493 和 0.0148，使用 $R_{123}G_{123}+A$ 比只使用 R_1 提高了 0.1068、0.0575 和 0.016。

3.2.2 不同的分类器对性能的影响

从表 5~8 中可以看出，在固定输入和模型融合方法，只对比分类器的情况下，随机森林在 Top25000 的查准率和查全率上的表现好于 GBDT，GBDT 在 AUC 上的表现好于随机森林。由于本文的实验更加关注查准率，可以认为随机森林的效果略

好于 GBDT。无论使用哪种分类器，将月粒度和日粒度的特征同时用来训练模型的效果均好于只使用一种时间粒度训练模型。

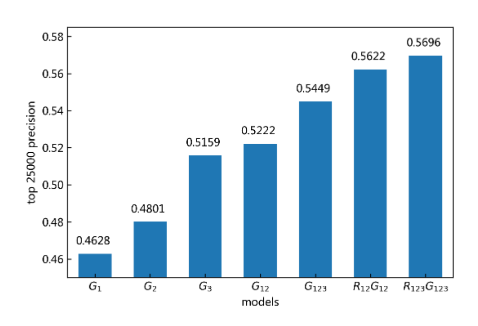


图 5　模型数量对性能的影响

3.2.3 不同的模型融合方法对性能的影响

从表 7~10 中可以看出，在本文数据集中使用平均法比使用 Stacking 融合的模型有更好的性能。同样地，无论使用哪种模型融合方法，将月粒度和日粒度的特征同时用来训练模型的效果均好于只使用一种时间粒度训练模型。

4　结束语

传统离网预测模型只选取单一时间粒度特征形成数据集并训练，这类方法只考虑了模型对性能的影响，没有 从数据的角度考虑，本文提出将不同时间粒度的特征融合，形成三组不同的数据集，在这些数据集上使用随机森林和 GBDT 训练，通过模型融合得到最终结果。实验证明，这种方法将模型的查准率从 0.4628~0.5147 提高到了 0.5696。多粒度时序特征的本质是在原有特征的基础上做特征工程，这种特征工程的方法只需要根据不同的时间粒度聚合特征，操作简单容易实现。本文验证了多粒度时序特征在离网预测任务中效果很好，并在此基础上提出了一个基于多粒度时序特征的离网预测框架。

目前本文只考虑融合月和日两种时间粒度，在进一步的研究中将会加入更多的时间粒度，同时尝试更复杂的融合方法，例如构建多层融合模型以期进一步提高模型性能。

参考文献：

[1] Coussement K, Van den Poel D. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques [J]. Expert Systems with Applications, 2008, 34 (1): 313-327.

[2] Verbeke W, Dejaeger K, Martens D, et al. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach [J]. European Journal of Operational Research, 2012, 218 (1): 211-229.

[3] Reinartz W J, Kumar V. The impact of customer relationship characteristics on profitable lifetime duration [J]. Journal of marketing, 2003, 67 (1): 77-99.

[4] Guyon I, Lemaire V, Boullé M, et al. Design and analysis of the KDD cup 2009: fast scoring on a large orange customer database [J]. ACM SIGKDD Explorations Newsletter, 2010, 11 (2): 68-76.

[5] Yu H F, Lo H Y, Hsieh H P, et al. Feature Engineering and classifier ensemble for KDD Cup 2010 [C]// Proc of JMLR: Workshop and Conference Proceedings. 2010: 1-16

[6] Umayaparvathi V, Iyakutti K. Automated feature selection and churn prediction using deep learning models [J]. International Research Journal of Engineering and Technology, 2017, 4 (3): 1846-1854.

[7] Castanedo F, Valverde G, Zaratiegui J, et al. Using deep learning to predict customer churn in a mobile telecommunication network [DB/OL]. http://www. wiseathena. com/pdf/wa_dl. pdf.

[8] Wangperawong A, Brun C, Laudy O, et al. Churn analysis using deep convolutional neural networks and autoencoders [J]. arXiv preprint arXiv: 1604. 05377, 2016.

[9] Zaratiegui J, Montoro A, Castanedo F. Performing highly accurate predictions through convolutional networks for actual telecommunication challenges [J]. arXiv preprint arXiv: 1511. 04906, 2015.

[10] Stripling E, Van den Broucke S, Antonio K, et al. Profit maximizing logistic regression modeling for customer churn prediction [C]// Proc of IEEE International Conference on Data Science and Advanced Analytics. 2015: 1-10.

[11] Lu Ning, Lin Hua, Lu Jie, et al. A customer churn prediction model in telecom industry using boosting [J]. IEEE Trans on Industrial Informatics, 2014, 10 (2): 1659-1665.

[12] Owczarczuk M. Churn models for prepaid customers in the cellular telecommunication industry using large data marts [J]. Expert Systems with Applications, 2010, 37 (6): 4710-4712.

[13] Zhao Xi, Shi Yong, Lee Jongwon, et al. Customer churn prediction based on feature clustering and nonparallel support vector machine [J]. International Journal of Information Technology & Decision Making, 2014, 13 (05): 1013-1027.

[14] Shaaban E, Helmy Y, Khedr A, et al. A proposed churn prediction model [J]. International Journal of Engineering Research and Applications, 2012, 2 (4): 693-697.

[15] Abbasimehr H, Setak M, Tarokh M J. A comparative assessment of the performance of ensemble learning in customer churn prediction [J]. Internal Arab Journal Information Technology, 2014, 11 (6): 599-606.

[16] Binti Oseman, K, Haris N A, bin Abu Bakar F. Data mining in churn analysis model for telecommunication industry [J]. Journal of Statistical Modeling and Analytics Vol, 2010, 1 (19-27) .

[17] Kirui C, Hong Li, Cheruiyot W, et al. Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining [J]. International Journal of Computer Science Issues, 2013, 10 (2): 1694-0784.

[18] Idris A, Khan A. Ensemble based efficient churn prediction model for Telecom [C]// Proc of the 12th International Conference on Frontiers of Information Technology. 2014: 238-244.

[19] Coussement K, Van den Poel D. Churn prediction in subscription services:

an application of support vector machines while comparing two parameter-selection techniques [J]. Expert Systems with Applications, 2008, 34 (1): 313-327.
[20] Verbeke W, Martens D, Mues C, et al. Building comprehensible customer churn prediction models with advanced rule induction techniques [J]. Expert Systems with Applications, 2011, 38 (3): 2354-2364.
[21] Kim N, Jung K H, Kim Y S, et al. Uniformly subsampled ensemble (USE) for churn management: Theory and implementation [J]. Expert Systems with Applications, 2012, 39 (15): 11839-11845.
[22] Wei C P, Chiu I T. Turning telecommunications call details to churn prediction: a data mining approach [J]. Expert Systems with Applications, 2002, 23 (2): 103-112.
[23] Zhang Junbo, Zheng Yu, Qi Dekang. Deep spatio-temporal residual networks for citywide crowd flows prediction [C]// Proc of the 31st AAAI Conference on Artificial Intelligence. 2017.
[24] Breiman L. Random forests [J]. Machine Learning, 2001, 45 (1): 5-32.
[25] De Groot D. S. Churn prediction in telecommunication [D]. Delft: Technische University Delft, 2017.